DOCUMENT RESUME

ED 441 441 IR 057 712

AUTHOR	Gorman, G. E.
TITLE	Collecting Data Sensibly in Information Settings.
PUB DATE	1999-08-00
NOTE	8p.; In: IFLA Council and General Conference. Conference Programme and Proceedings (65th, Bangkok, Thailand, August 20-28, 1999); see IR 057 674.
AVAILABLE FROM	For full text: http://www.ifla.org/IV/ifla65/papers/004-120e.htm.
PUB TYPE	Opinion Papers (120) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Data Collection; Foreign Countries; Information Management; Library Materials; *Library Services; *Library Statistics; *Measurement Techniques; Users (Information)
IDENTIFIERS	*Output Measurements; *Variables

ABSTRACT

Libraries and information agencies depend heavily on statistics to describe their services, evaluate their activities, and measure their performance. In the data gathering on which statistical analysis depends, there are always assumptions and uncontrolled variables that interfere with the purity and objectivity of the data, and therefore contaminate the analysis and interpretation of that data. This paper highlights some of these variables in order to alert information managers to the pitfalls of data collection and to encourage them to develop means of controlling data so that they can use statistics more effectively. Topics addressed include: (1) whether users can be counted meaningfully; (2) the value of counting holdings; (3) counting inquiries as a substitute for counting users or holdings; (4) problems with external stakeholders; and (5) suggestions for enhancements to the standard statistical measures employed in the information sector. (Contains 12 references.) (Author/MES)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 441 441

IFLANET Search Contacts
International Federation of Library Associations and Institutions
Annual Conference

65th IFLA Council and General Conference

Bangkok, Thailand, August 20 - August 28, 1999

Conference Proceedings

Code Number: 004-120-E
Division Number: VI
Professional Group: Statistics
Joint Meeting with: -
Meeting Number: 120
Simultaneous Interpretation: *No*

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

A.L. Van Wesermael

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Collecting Data Sensibly in Information Settings [1]

G. E. Gorman
School of Communications and Information Management
Victoria University of Wellington
Wellington 6001, New Zealand
E-mail: gary.gorman@vuw.ac.nz

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Abstract

Libraries and information agencies depend heavily on 'statistics' to describe their services, evaluate their activities and measure their performance. In the data gathering on which statistical analysis depends, there are always assumptions and uncontrolled variables that interfere with the purity and objectivity of the data, and therefore contaminate the analysis and interpretation of that data. This paper highlights some of these variables in an attempt to alert information managers to the pitfalls of data collection and to encourage them to develop means of controlling data so that they can use statistics more effectively.

Paper

The Focus on Statistics

A major management activity in libraries at the end of the 20th century is data collection and the production of statistics. For the most part library managers is that more is better - more data will lead to more useful information, which will produce more informed decisions and, therefore, a more adequately managed service. The underlying assumption is that data about library activities can be transformed into useful information, and that the information will become management knowledge.

It is understandable, then, that data collection is viewed by many as the most basic activity in

IR 057712

BEST COPY AVAILABLE

the management process. But it is less understandable that managers tend on the whole to view the data collection ? interpretation ? application process as something to be accepted without question, and that many simply apply this model time after time without considering whether there might be a better way to collect and utilise data. To a social scientist whose primary interest is research methodologies and whose primary employment is teaching Library and Information Science (LIS) students about research, it is a worrying state pf affairs that has been with us for several decades.

The principal purpose of this address is to suggest that information professionals might profitably consider using not only quantitative but also qualitative data collection and analysis methods in order to achieve greater reliability and deeper meaning in their investigations. A secondary purpose is to highlight some of the dangers inherent in the unquestioning acceptance of the data collection and interpretation process.

Quantitative variables tell how much or how many of a particular characteristic or attribute are present in an object, event, person or phenomenon. One example is the number of computers available for students to use on a campus. Qualitative variables classify an object, event, person or phenomenon into categories with respect to the attributes by which they differ. For example, the language of publication of a given journal title may be English, French, Hebrew or Spanish .

By looking beyond 'how much' and 'how many' to the attributes of the people, things and activities being counted, librarians cannot help but have a more useful understanding of their organisations and their work.

This is not a new concern, nor are the solutions offered in this paper unique; but no matter what has been said in the past, the problem remains and it seems worthwhile to rehearse the realities yet again. One of my Victoria University colleagues, Rowena Cullen, has queried the value of relying on quantitative data alone in the context of her research on performance measurement. Thus, discussing the work by Pratt and Altman and by the Library and Information Statistics Unit (LISU) , she wonders about the reliability of library statistics alone as a reliable measure of library activity, and especially whether such data can enable much correlation between inputs and outputs. 'In particular substantial issues of user satisfaction with their library/information services are touched on by only a small percentage of studies included here, although the authors comment in several places that further analysis is possible and indeed desirable.'

Cullen goes on in her paper to demonstrate that a library is a social construct and that, therefore, performance measurement is also a social construct. This then means that we need to be looking at a matrix incorporating values, focus and purpose -- three axes essential in understanding the library as social construct. In my view the social construct is a means pf viewing libraries and information organisations, and when we are in the realm of social constructs non-quantitative methods of data collection and analysis become more meaningful. This is especially so in three areas: library users, collections and services (or enquiries); each of these is discussed in turn.

Can Users Be Counted Meaningfully?

Data collection is built on the assumption that it is possible to arrive at a fair representation of the objects/population under investigation. In a library such an assumption must be questioned when it is applied to the user population of a particular library service. As an example, suppose we are interested in the number of people using the library. How do we count or measure this? One simple way is just to count those entering or leaving the building, either mechanically or manually. And many libraries do precisely this - how many annual reports proudly boast that 'in 199X the library was visited by XXXX users'? But what does this tell us? Were these casual users, serious scholars making intensive use of sophisticated search services, students looking for materials on a reading list, elderly people using the library as a social centre,

parents taking advantage of programs for their children? In other words, counting people tells us very little, as it does not specify the various categories of users and thus the demands they are likely to make on the service. This is a prime example of data being unable to generate meaningful statistics or information of any value because it adopts such a crude perception of the user population. The basic assumption is flawed, the data are flawed, and thus the interpretation must be equally flawed.

What we really want is a profile of actual users of a library - who they are, what they expect when they enter the library, what use they make of the facilities and services, what they think of the facilities and services, why they might choose to access the library electronically or in person, etc. None of this very useful data is obtainable by a simple counting of users. Furthermore, no amount of counting, even the most sophisticated and detailed survey of users, can tell us anything about the potential users or the non-users, yet surely this sort of information is what managers really want - they want, or at least they need, to know about the potential market for their services so that they can produce a management plan for tapping into this reserve. Even in such a library-conscious country as Australia, with more than 60% of the population using public libraries (whatever 'using' might mean), there is a very large non-user population that we need to draw into our service. Even the most accomplished researcher will tell you that collecting this more useful, and therefore more sophisticated, data on users and non-users is fraught with difficulties, and that it is a time-consuming and expensive proposition. But as yet there is no substitute for this.

What Is the Value of Counting Holdings?

If I am correct in querying the value of counting users, as distinct from counting distinctive cohorts of users and determining their assessment of particular services, is it possible to shift our focus from people to objects - specifically, holdings (however defined)?

It is almost a Biblical truth in libraries, especially since the 'good old days' when the Clapp-Jordan formula was in its ascendance, that counting the size of a book collection gives us data that are meaningful in quantitative and qualitative terms. Again, almost every annual report states that ' the size of the collections has now reached X number of books, Y number of current serial subscriptions, Z number of electronic resources'. But what is the relationship between the size or quantity of a collection and its quality? This is a question that invariably frustrates statisticians, because it calls into question the value of the statistical enterprise. But, as with users, we as information professionals must be primarily interested in values and meanings, whether we are looking at users or collections.

Assuming that there is a relationship between quantity and quality, and I certainly do not make this assumption, it is necessary to question the value of data on holdings as assuming any relationship between size and level of service. To compensate for this, many libraries count loans or number of uses of books, reference materials, journals, CD-ROMs, etc. However, any counting of loans or uses is easily skewed by an unusual and out-of-the-ordinary use by a scholar working on a one-off project, by a borrower with a passing fancy in a particular topic, etc. It might also be questioned, in an era of entrepreneurial focus and value-added service, whether loans or uses of library materials is a valid indicator of much.

It is possible, of course, to enhance data on loans and uses by introducing some sort of quality indicators into collections. This tends to mean a ranking system of some sort, usually one which matches items in a local collection against some external measure. In New Zealand this might mean that Wellington City Library, for instance, puts a higher value on its items which are also held in the New York Public Library. Or a university library might rank publications of prestigious academic publishers and U.N. agencies above popular novels or local government publications. But is a library or information service meant to be responsive to the needs of a specific local community, or is it in the business of measuring itself against national or international criteria? That is, for the Wellington City Library it may be - perhaps even ought to be - that materials unique to its collection, and not those also in the New York Public

Library, are most relevant, to local user requirements.

In other words, to count holdings, whether of books or any other medium, is not a measure of demand; to count uses is not a measure of the level or quality of use, only that items have been taken off the shelves or accessed electronically, perhaps because nothing 'better' is available. But is this begging the question? For too many librarians levels of demand for services is not a significant issue, whereas size of holdings or number of uses is.

Are Enquiries a Substitute for Counting Users or Holdings?

Counting users and collections may give us some data, albeit of limited value, but one feels compelled to reiterate that too many library services hide behind such raw figures, and rely on these as a substitute for meaningful data analysis. One alternative adopted by some institutions is to count user enquiries (of staff, of electronic systems or other question-answering modes). Some excellent examples of this can be found in Libraries in the Workplace, one of those excellent reports generated by David Spiller and LISU associates: How many searches (end user and mediated) do you estimate were made from the library/information centre? How many enquiries do you estimate were answered from the library/information centre?

When asking about the number of enquiries, libraries tend to record the number of queries over a given period of time, or observe perceived user interactions with inanimate information resources. As always with data collection, it is relatively easy for the data to be skewed or distorted by

the recorder - usually a member of the library staff, who may well feel threatened by the procedure and therefore may pad the figures to make the enquiry service look busier than it actually is. For example, a staff member may intentionally alter the figures to include a larger number of queries than were actually made; or, more typically, a simple directional request may be treated as a query, when in reality the staff are to be counting only information requests.

As with stock circulation questions, we want to know something about the level of queries. Are all queries equal? No. Do some take more time and greater effort? Of course. So why not ask questions that generate data about the time and amount of detail provided in response to queries?

Consider how much richer the data might be if the following question were asked: Of the total number of enquiries answered from the library/information centre, what percentage do you estimate took the following amount of time: then state a range of times, from 1 minute to 10 minutes, etc.? Or what about asking for information on the type of query: Was it recreational, informational, research-oriented? Would questions such as these give us better insights into the nature, depth and quality of service being provided?

As enquiry systems become increasingly automated, it is relatively straightforward to build recording mechanisms into electronic systems, allowing for retrieval of data on length of queries, amount of data retrieved, etc. While we are on the topic of automated systems, there is also the matter of counting user-machine interactions. Here it is more difficult to corrupt the data, or at least easier to eliminate non-informational queries from the count.

At the other extreme is collecting data about perceived user interactions, which is notoriously unreliable because of its dependence on detached, unobtrusive observation. This method of data collection is particularly open to bias, especially in a library setting where cheap labour (i.e., student observers) is employed. This can lead to the '...selective recording of observational data. Certain objects and relations may more likely be recorded by observers with different interests, biases and backgrounds.' In other words, observation skills are essential, and to the extent that these are flawed, the data will be flawed. Allan Kellehear's excellent work on observation contains a number of caveats about this data collection

techniques, all of which can be summarised as follows: the observer must be skilled in observing and must never impute any motives to the observed interaction or behaviour. In an information setting the natural tendency is to assume that an interaction is in some way task-related (a user is seeking information for a specific purpose), and this is to impute a motive that may well not exist.

The Problem with the Stakeholders...

Of course, one problem with academics who plead for richer data collection techniques is that, as all practitioners know, we live in ivory towers far removed from 'the real world'. Indeed. And it has to be recognised that in that 'real world' the stakeholders for whom much data collection and analysis are undertaken simply do not want to have much detail, do not want to have to think about data, and just want a simple table that shows how institution X is better than institution Y ('better' meaning a bigger budget, more reference transactions, larger bookstock, etc.). That is, we need to recognise that data collection is driven to a considerable extent by those to whom the practitioners are accountable, and those to whom we are accountable as often as not have bean-counting mentalities.

Whether the stakeholders are administrative, managerial, political or financial, it is important to recognise that they have the power to dictate what data we collect, how the data are used and how they are presented. Every library or information service is accountable to someone else in that they depend on that someone for funding, for their very raison d'etre. The 'someone else' needs to understand the information needs of libraries. If external stakeholders are allowed to dictate data collection needs and presentation standards, then it is totally realistic to expect them to structure these for their own interests rather than for those of the library - and why shouldn't they?

The increasing sophistication of automated library systems, and the greater ease with which numeric data can be collected - on users, on collections, on expenditures, on transactions - means that we are becoming more wedded than ever to simple quantification as a means of evaluation. As this occurs, stakeholders believe more rigidly that data can be collected most simply by means of a keystroke here, a command there. Consequently, it becomes less likely that we can break out of the number-crunching mold, because our controllers continue to see this as the most effective way of evaluating our services. Also, it must be admitted, software that ought to aid in the analysis of qualitative data (which are not simple to analyse) simply lack the user-friendliness and ease of interpretation required in data analysis. Despite the positive assessments by evaluators such as Miles and Huberman of qualitative data analysis software, one remains sceptical of most commercially-available packages. Computer software, after all, uses technical processing methods for qualitative data that intrinsically are more suited to other, more time-consuming methods.

There is a significant distinction to be made between efficiency (the lowest per unit cost of something) and effectiveness (successful accomplishment of a task or mission). Our stakeholders almost invariably are efficiency-driven, and the technology that enhances data collection and analysis certainly enhances efficiency (and only efficiency). We information professionals, in contrast, are members of a service industry in which successful accomplishment of our mission - effectiveness - should be paramount.

What Can Be Done?

There are a number of implications in the preceding discussion about what we might do to change the situation from number-driven, efficiency conscious data collection and analysis to more context-sensitive, sense-making collecting and analytical techniques. All of these are offered not as alternatives, but as enhancements to, the standard statistical measures employed universally in the information sector.

- Look seriously at the genuine shortcomings of quantitative data collection and analysis methods and seek to incorporate qualitative methods that permit deeper understanding of library users, collections and services.
- Focus less on users as a genus, more on specific categories of users and profiles of their wants and needs.
- Focus less on numerical aspects of collections and more on acceptable indicators of collection quality.
- Focus less on simple user enquiries and more on the nature and level of these enquiries.
- Employ qualitative data collection methods in full awareness of the problems associated with achieving value-free use of these methods.
- Foster an awareness among stakeholders that efficiency and effectiveness are not equivalent concepts, and that effectiveness in the information sector is a greater good than efficiency.
- Work with software developers in creating qualitative data software that are more acceptable in terms of user friendliness and analytical capabilities.

Conclusion

A recent paper by Dole and Hurych discusses 'new measurements' for library evaluation, especially with regard to electronic resources. The authors provide an excellent review of conventional measures and also offers clear insights into current developments. It is heartening to see that use-based measures are being considered, but depressing that these form a very small component of conventional cost-, time- and transaction-based measures. If this is the future of data collection in libraries, then I am not convinced that we will see much improvement in what I regard as a less-than-adequate situation.

More promising is some work being encouraged by the U.S.-based Coalition for Networked Information (http://www.cni.org), and in particular by Charles McClure. In Assessing the Academic Networked Environment: Strategies and Options he and Cynthia Lopata present a network assessment manual that is largely qualitative in its approach, and that makes a strong case for using qualitative methods in assessing academic networks. However, this seems not to have been greeted with universal acclaim, and certainly has not made much of an impact on the data-collecting community.

In the final analysis what we are arguing for is a greater awareness among library professionals that meaningful data are contextual and that meaning depends on interpretation, that they are derived from variables that are complex and difficult to measure, that understanding is an inductive process. This differs from, but is not necessarily in conflict with, the traditional quantitative approach of the statistician that assumes the possibility of identifying and measuring variables in a relatively straightforward manner, that norms and consensus can be derived from the data by deduction. Both have their place in information work, but please let us not emphasise one at the expense of the other - or rather continue to emphasise one (quantitative) at the expense of the other (qualitative).

Remember the classic work by Webb et al. on unobtrusive measures, in which Chapter 8 contained a statistician's impassioned plea for researchers to use 'all available weapons of attack'? More than 30 years later, it is high time that information professionals heed the call and look beyond their numbers to sources of potentially deeper meaning.

References

1. Hafner, A.W. (1998). Descriptive Statistical Techniques for Librarians. 2nd ed. Chicago: American Library Association.

2. Pratt, A., and Altmann, E. (1997). 'Live by the Numbers, Die by the Numbers.' Library Journal April 15: 48-49.

3. England L., and Sumsion, J. (1995). Perspectives of Public Library Use: A Compendium of Survey Information. Loughborough: Library and Information Statistics Unit.

4. Cullen, R. (1998). "Does Performance Measurement Improve Organisational Effectiveness? A Post-modern Analysis.' In Proceedings of the 2nd Northumbria International Conference on Performance Measurement in Libraries and Information Services Held at Longhirst Hall, Northumberland, England, 7-11 September, 1997, 3-20. Newcastle upon Tyne: Information North.

5. Clapp, V.W., and Jordan, R.T. (1965). 'Quantitative Criteria for Adequacy of Academic Library Collections.' College and Research Libraries 26: 371-380.

6. Spiller, D.; Creaser, C.; and Murphy, A. (1998). Libraries in the Workplace. LISU Occasional Papers, 20. Loughborough: Loughborough University of Technology, Library and Information Statistics Unit.

7. Kellehear, A. (1993). The Unobtrusive Researcher: A Guide to Methods. St Leonards, NSW: Allen and Unwin.

8. Miles, M.B., and Huberman, A.M. (1994). Qualitative Data Analysis: An Expanded Sourcebook . 2nd ed. Thousand Oaks, CA: Sage Publications.

9. Dole, W.V., and Hurych, J. M. (1999). 'New Measurements for the Next Millennium: Evaluating Libraries in the Electronic Age.' paper prepared for CoLIS3: The Third International Conference on Conception of Library and Information Science, Dubrovnik, Croatia, 23-27 May.

10. McClure, C.R., and Lopata, C. (1996). Assessing the Academic Networked Environment: Strategies and Options. February 1996.

11. Gorman, G.E., and Clayton, P.R. (1997). Qualitative Research for the Information Professional: A Practical Handbook. London: Library Association Publishing.

12. Webb, E., et al. (1966). Unobtrusive Measures: Non-Reactive Research in the Social Sciences. Chicago: Rand-McNally.

Latest Revision: *May 27, 1999*

Copyright © 1995-1999
International Federation of Library Associations and Institutions
www.ifla.org